wwPDB EM Validation Summary Report

May 21, 2022 – 07:18 am BST

PDB ID	:	7R5V
EMDB ID	:	EMD-14341
Title	:	Structure of the human CCAN CENP-A alpha-satellite complex
Authors	:	Yatskevich, S.; Muir, K.W.; Bellini, D.; Zhang, Z.; Yang, J.; Tischer, T.; Predin, M.; Dendooven, T.; McLaughlin, S.H.; Barford, D.
Deposited on	:	2022-02-11
Resolution	:	4.55 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

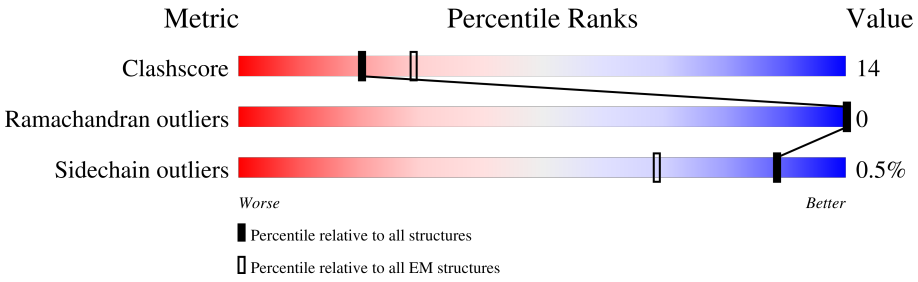
EMDB validation analysis	:	0.0.1.dev8
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 4.55 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



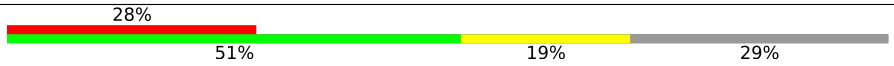
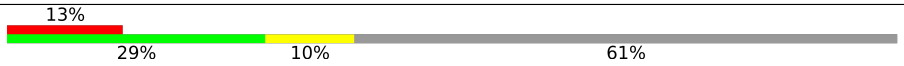
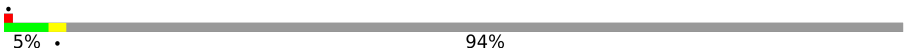
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	H	247	34% 16% 50%
2	I	756	30% 13% 57%
3	J	171	16% 81%
4	K	269	36% 11% 53%
5	L	344	62% 25% 14%
6	M	180	57% 39%
7	N	339	60% 28% 12%
8	O	300	47% 23% 30%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	P	288	
10	Q	268	
11	U	418	
12	i	171	
13	R	177	

2 Entry composition (i)

There are 13 unique types of molecules in this entry. The entry contains 18507 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Centromere protein H.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	H	124	1013	636	176	195	6	0	0

- Molecule 2 is a protein called Centromere protein I.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	I	328	2686	1777	421	472	16	0	0

- Molecule 3 is a DNA chain called DNA (171-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	J	32	658	313	116	197	32	0	0

- Molecule 4 is a protein called Centromere protein K.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	K	126	1026	637	170	211	8	0	0

- Molecule 5 is a protein called Centromere protein L.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	L	297	2397	1563	392	428	14	0	0

- Molecule 6 is a protein called Centromere protein M.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	M	172	1325	839	236	243	7	0	0

- Molecule 7 is a protein called Centromere protein N.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	N	297	Total	C	N	O	S	0	0
			2452	1581	425	436	10		

- Molecule 8 is a protein called Centromere protein O.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	O	210	Total	C	N	O	S	0	0
			1642	1060	277	298	7		

- Molecule 9 is a protein called Centromere protein P.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	P	219	Total	C	N	O	S	0	0
			1732	1107	293	324	8		

- Molecule 10 is a protein called Centromere protein Q.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	Q	189	Total	C	N	O	S	0	0
			1506	943	253	299	11		

- Molecule 11 is a protein called Centromere protein U.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	U	165	Total	C	N	O	S	0	0
			1346	851	234	256	5		

- Molecule 12 is a DNA chain called DNA (171-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
12	i	31	Total	C	N	O	P	0	0
			632	300	120	181	31		

- Molecule 13 is a protein called Centromere protein R.

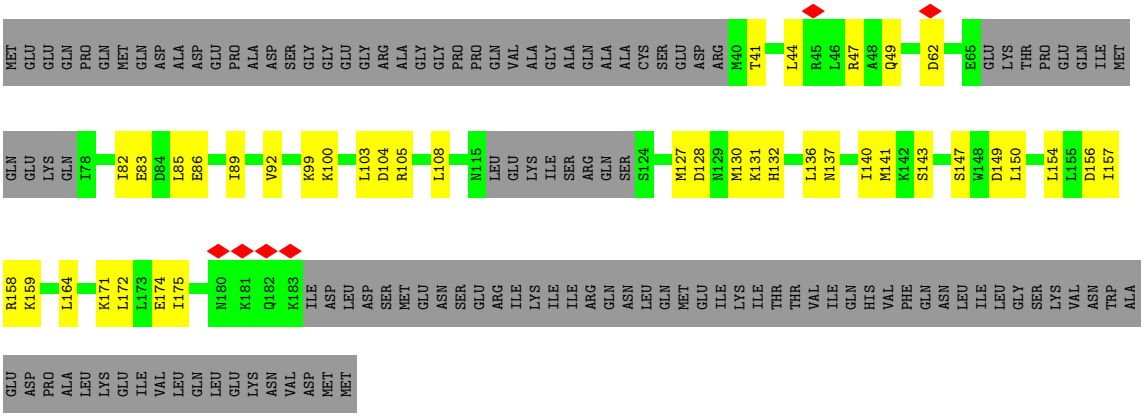
Mol	Chain	Residues	Atoms				AltConf	Trace
13	R	11	Total	C	N	O	0	0
			92	61	13	18		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

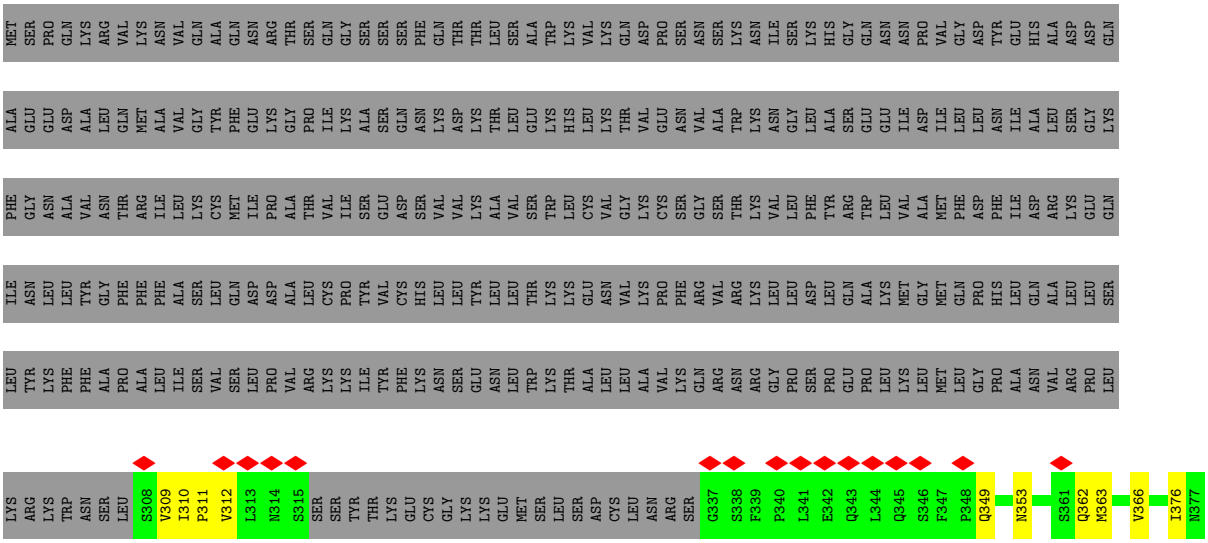
• Molecule 1: Centromere protein H

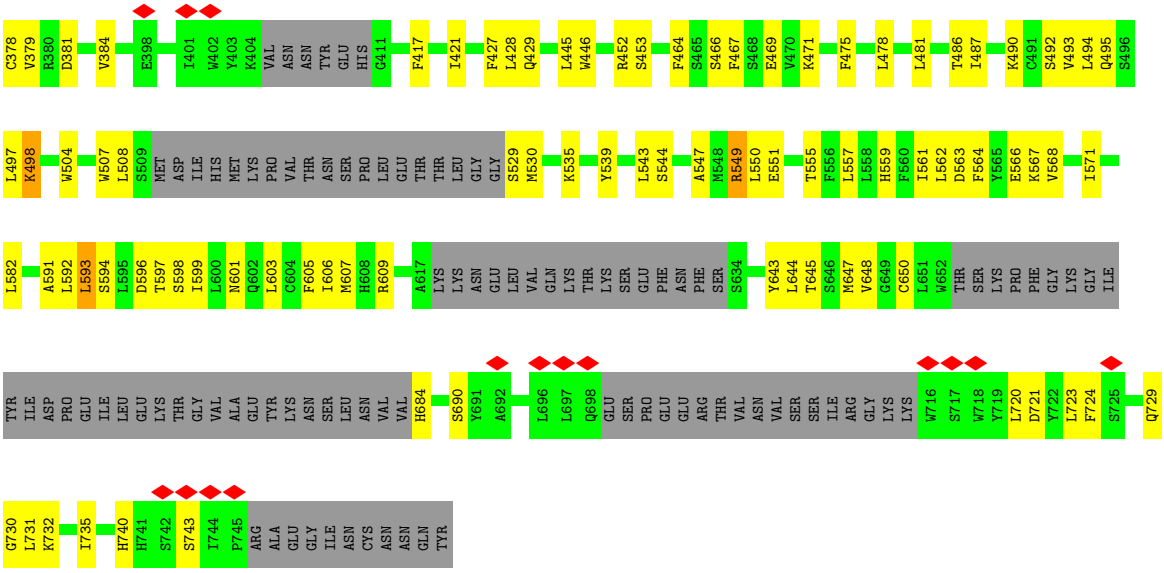
Chain H:

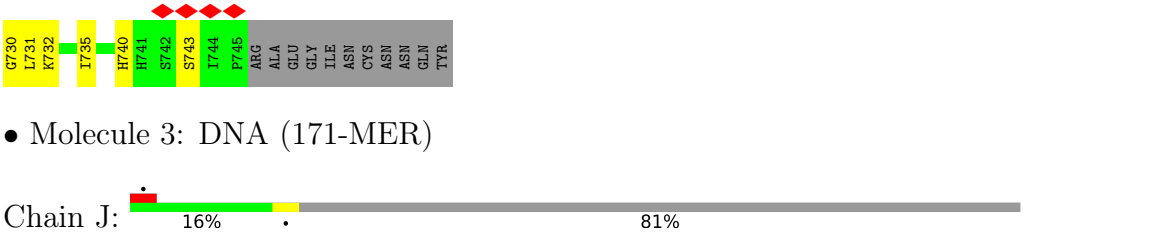


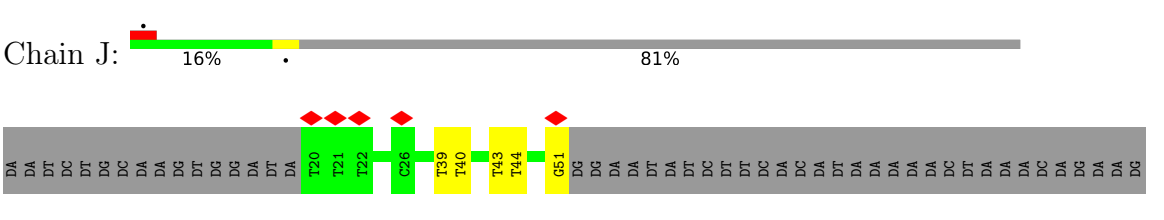
• Molecule 2: Centromere protein I

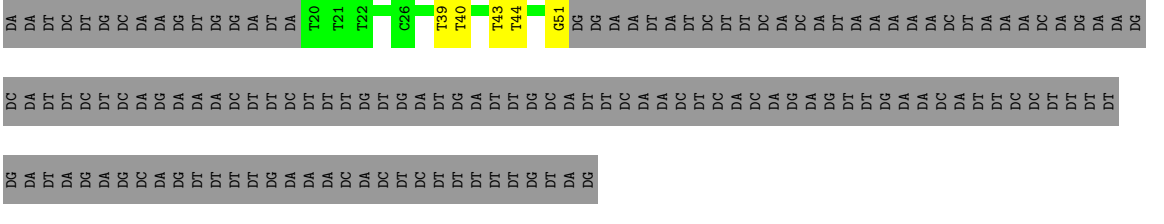
Chain I:

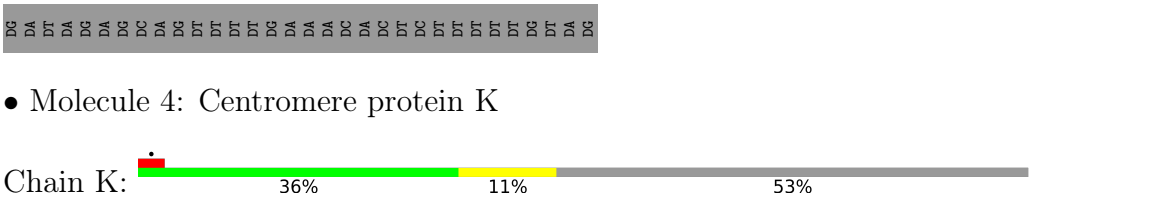


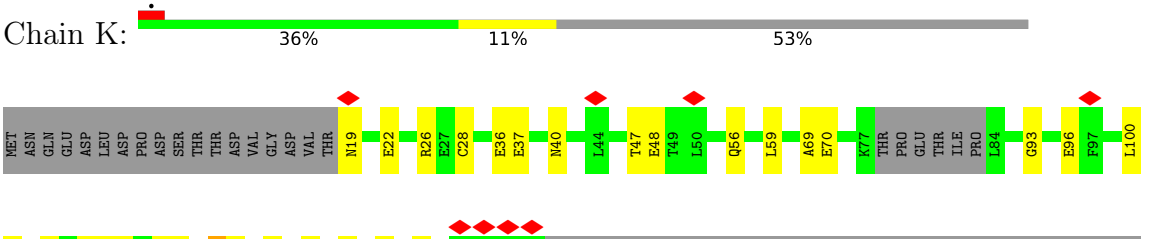


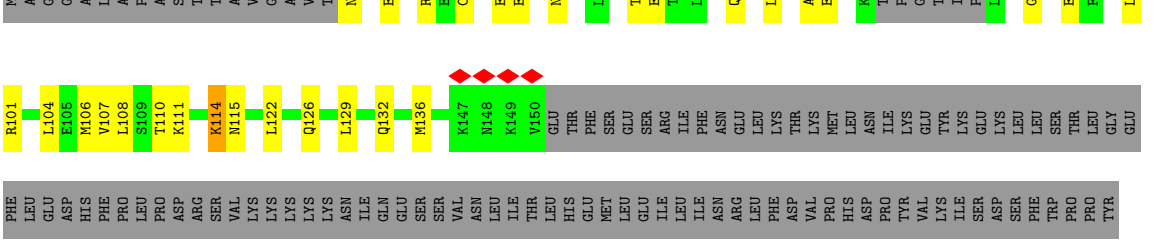


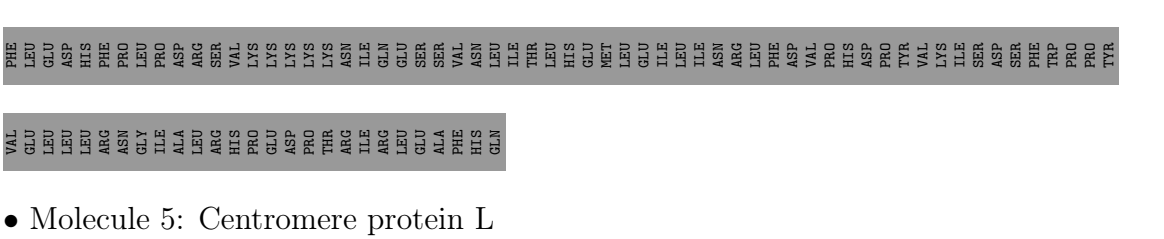


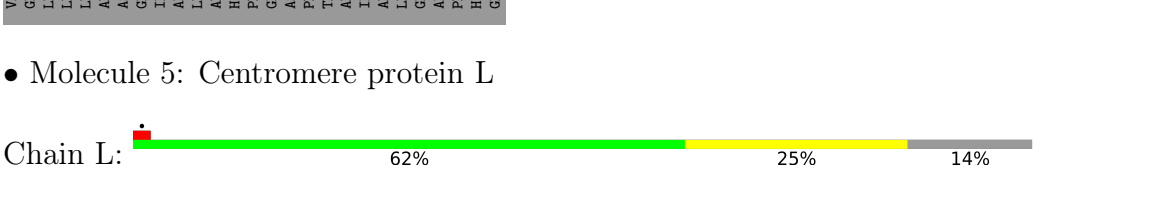












- Molecule 3: DNA (171-MER)

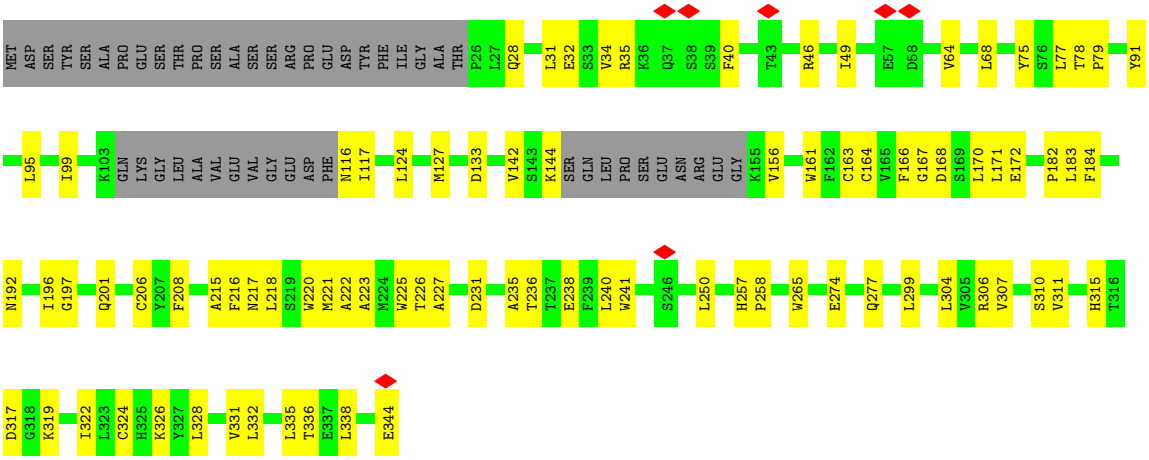
Chain J:

- Molecule 4: Centromere protein K

Chain K:

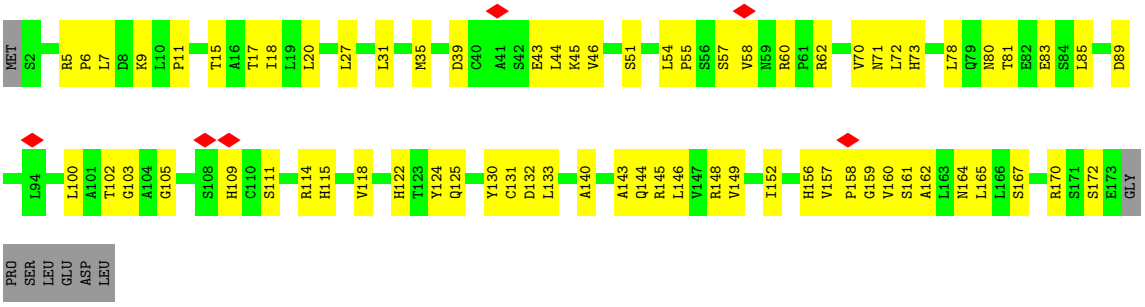
- Molecule 5: Centromere protein L

Chain L:



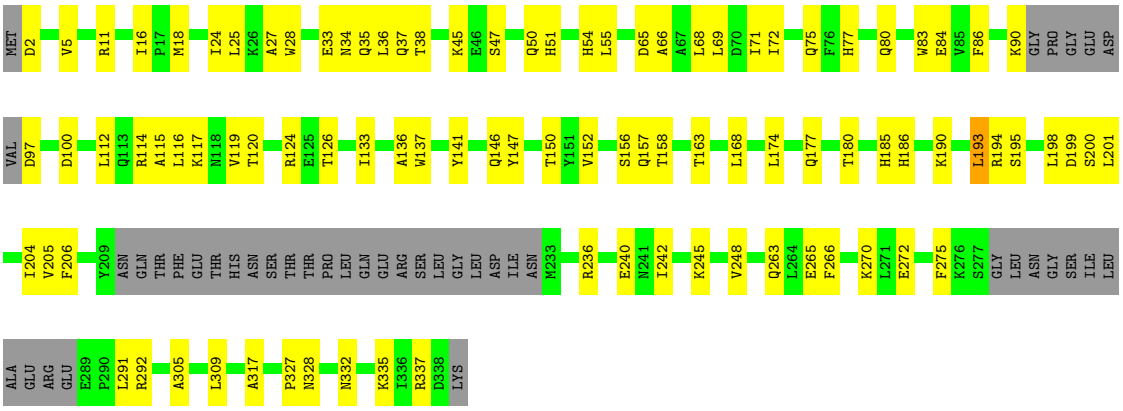
• Molecule 6: Centromere protein M

Chain M:



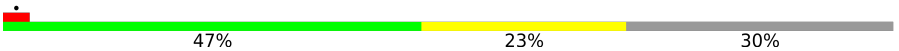
• Molecule 7: Centromere protein N

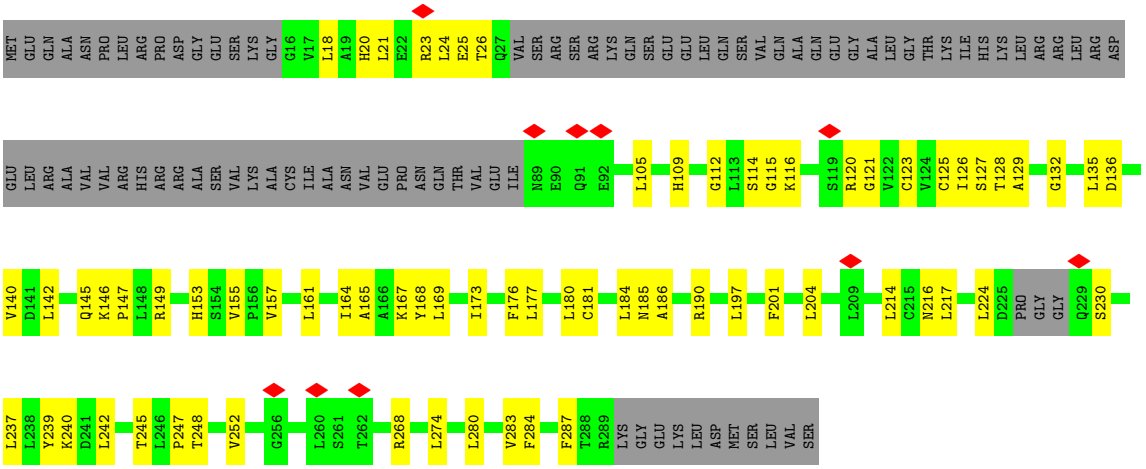
Chain N:



• Molecule 8: Centromere protein O

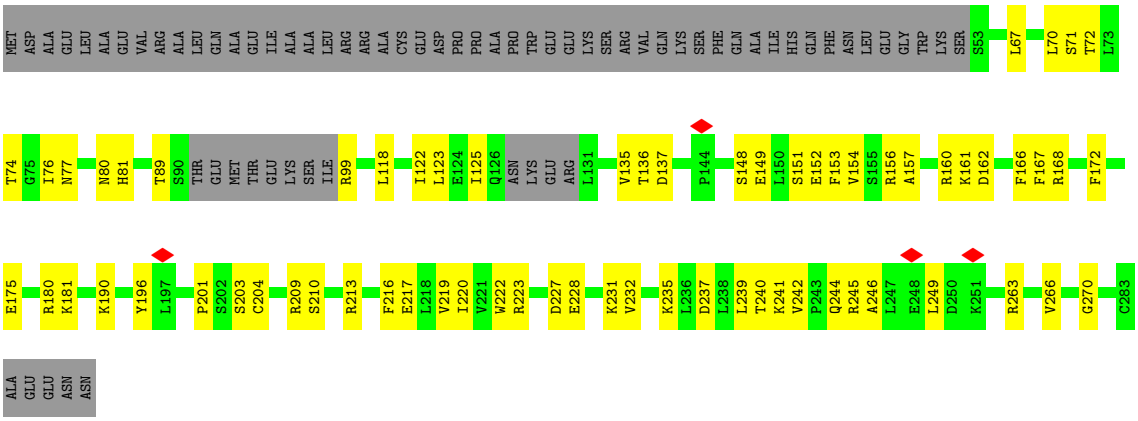
Chain O:





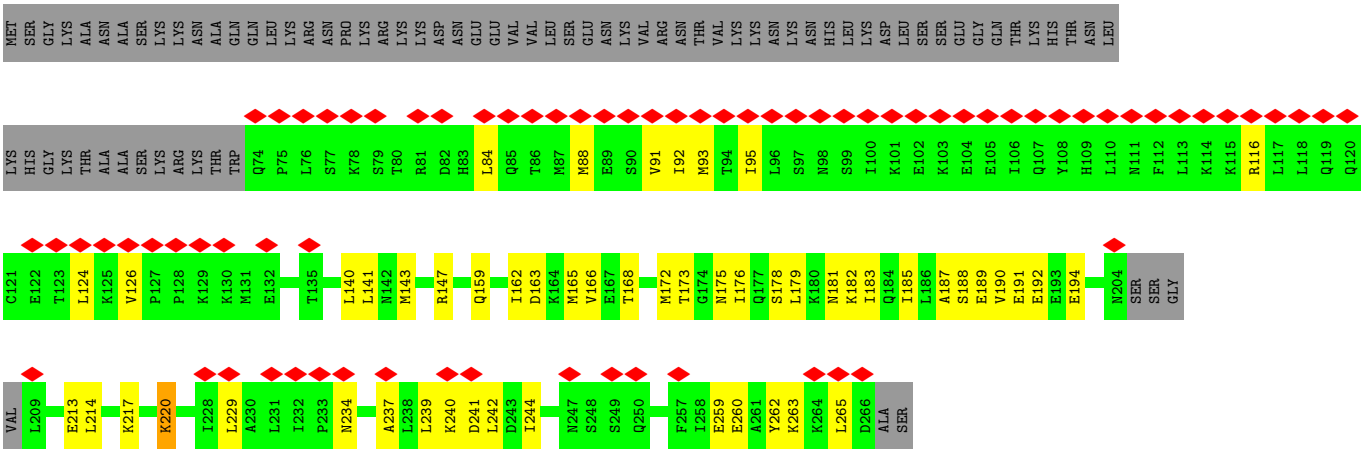
• Molecule 9: Centromere protein P

Chain P:



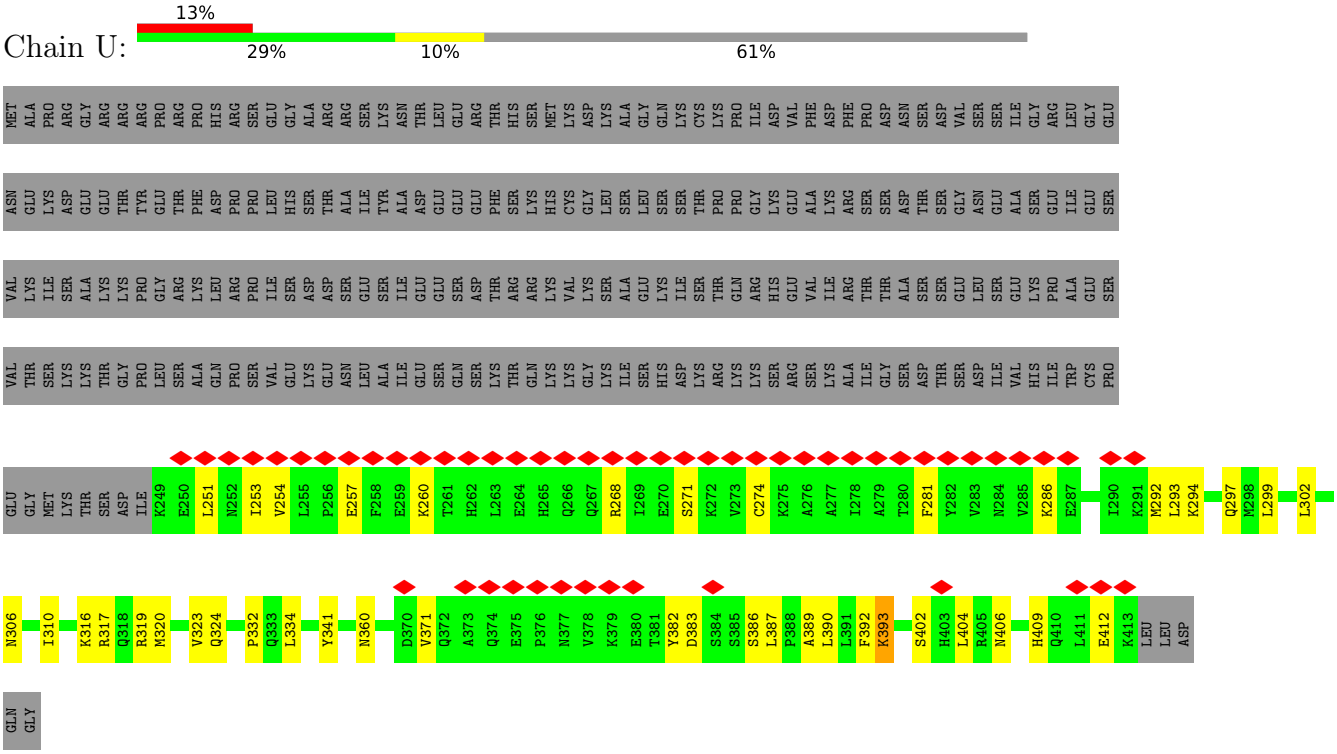
• Molecule 10: Centromere protein Q

Chain Q:



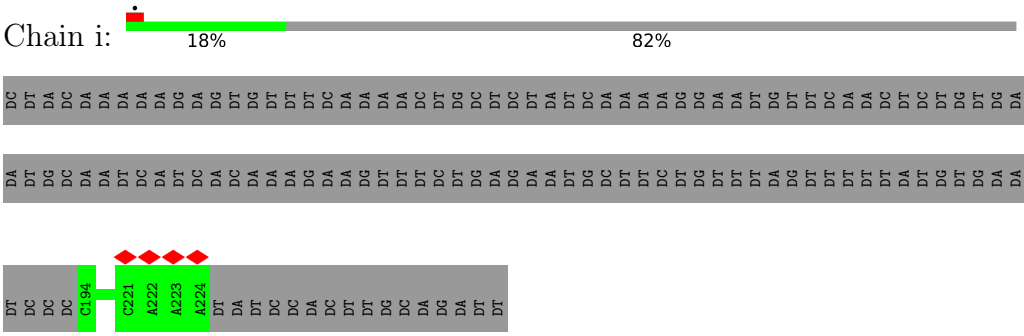
• Molecule 11: Centromere protein U

Chain U:



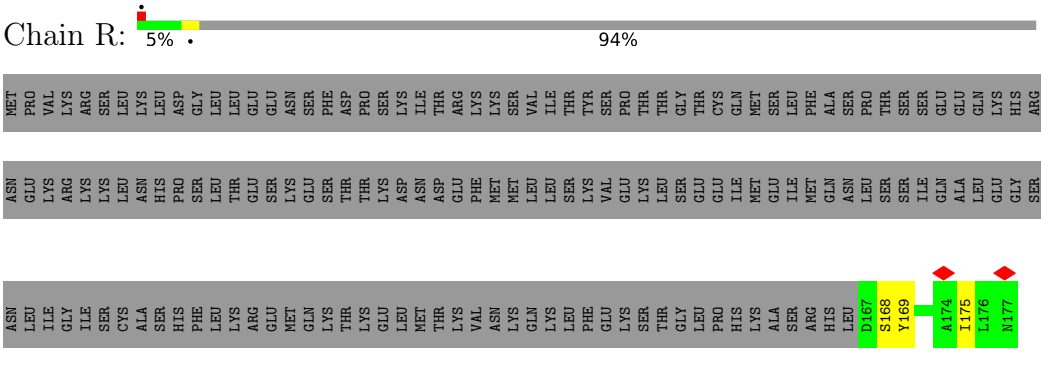
• Molecule 12: DNA (171-MER)

Chain i:



• Molecule 13: Centromere protein R

Chain R:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	123677	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	2600	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	1.415	Depositor
Minimum map value	-0.695	Depositor
Average map value	0.005	Depositor
Map value standard deviation	0.040	Depositor
Recommended contour level	0.23	Depositor
Map size (Å)	319.104, 319.104, 319.104	wwPDB
Map dimensions	384, 384, 384	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.83100003, 0.83100003, 0.83100003	Depositor

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	H	0.28	0/1015	0.49	0/1348
2	I	0.25	0/2764	0.43	1/3758 (0.0%)
3	J	0.49	0/736	0.92	0/1135
4	K	0.25	0/1032	0.43	0/1382
5	L	0.26	0/2462	0.43	0/3340
6	M	0.25	0/1347	0.46	0/1827
7	N	0.29	0/2505	0.44	0/3380
8	O	0.26	0/1678	0.47	0/2280
9	P	0.28	0/1763	0.46	0/2377
10	Q	0.26	0/1518	0.44	0/2036
11	U	0.25	0/1364	0.43	0/1831
12	i	0.47	0/709	0.84	0/1090
13	R	0.26	0/93	0.34	0/124
All	All	0.29	0/18986	0.50	1/25908 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	593	LEU	CA-CB-CG	5.34	127.58	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	H	1013	0	1071	40	0
2	I	2686	0	2619	80	0
3	J	658	0	363	4	0
4	K	1026	0	1025	27	0
5	L	2397	0	2395	68	0
6	M	1325	0	1370	54	0
7	N	2452	0	2485	70	0
8	O	1642	0	1616	72	0
9	P	1732	0	1715	48	0
10	Q	1506	0	1563	41	0
11	U	1346	0	1368	37	0
12	i	632	0	347	0	0
13	R	92	0	89	12	0
All	All	18507	0	18026	471	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 14.

The worst 5 of 471 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
8:O:173:ILE:CD1	13:R:168:SER:HB2	1.69	1.20
8:O:173:ILE:CD1	13:R:168:SER:CB	2.19	1.19
8:O:173:ILE:HD13	13:R:168:SER:HB2	1.25	1.09
8:O:173:ILE:HD12	13:R:168:SER:CB	1.94	0.97
2:I:729:GLN:HA	2:I:732:LYS:HE2	1.59	0.84

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	H	118/247 (48%)	118 (100%)	0	0	100	100
2	I	314/756 (42%)	304 (97%)	10 (3%)	0	100	100
4	K	122/269 (45%)	118 (97%)	4 (3%)	0	100	100
5	L	291/344 (85%)	282 (97%)	9 (3%)	0	100	100
6	M	170/180 (94%)	166 (98%)	4 (2%)	0	100	100
7	N	289/339 (85%)	277 (96%)	12 (4%)	0	100	100
8	O	204/300 (68%)	196 (96%)	8 (4%)	0	100	100
9	P	213/288 (74%)	209 (98%)	4 (2%)	0	100	100
10	Q	185/268 (69%)	179 (97%)	6 (3%)	0	100	100
11	U	163/418 (39%)	162 (99%)	1 (1%)	0	100	100
13	R	9/177 (5%)	9 (100%)	0	0	100	100
All	All	2078/3586 (58%)	2020 (97%)	58 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	114/224 (51%)	114 (100%)	0	100	100
2	I	304/691 (44%)	301 (99%)	3 (1%)	76	86
4	K	118/260 (45%)	117 (99%)	1 (1%)	81	89
5	L	267/306 (87%)	267 (100%)	0	100	100
6	M	151/158 (96%)	151 (100%)	0	100	100
7	N	270/311 (87%)	268 (99%)	2 (1%)	84	90
8	O	177/263 (67%)	177 (100%)	0	100	100
9	P	191/259 (74%)	190 (100%)	1 (0%)	88	93
10	Q	176/248 (71%)	175 (99%)	1 (1%)	86	92
11	U	149/379 (39%)	147 (99%)	2 (1%)	69	82

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
13	R	10/166 (6%)	10 (100%)	0	100 100
All	All	1927/3265 (59%)	1917 (100%)	10 (0%)	89 93

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
10	Q	220	LYS
11	U	360	ASN
11	U	393	LYS
4	K	114	LYS
7	N	126	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
7	N	186	HIS
8	O	185	ASN
8	O	212	ASN
4	K	40	ASN
2	I	314	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

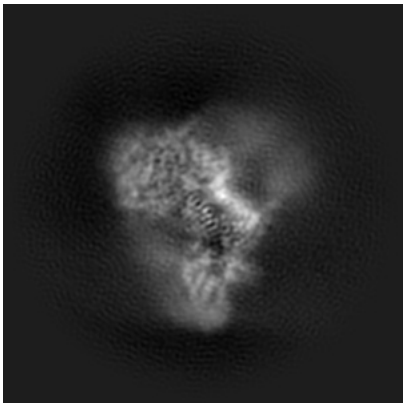
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-14341. These allow visual inspection of the internal detail of the map and identification of artifacts.

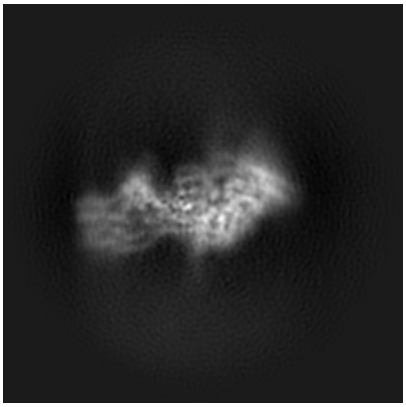
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

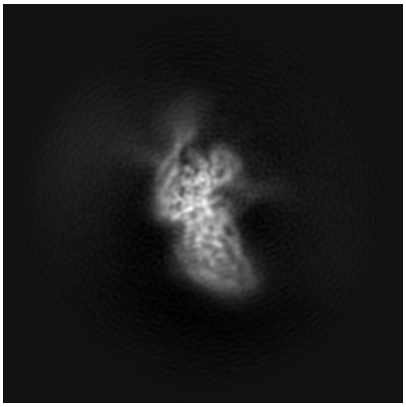
6.1.1 Primary map



X



Y

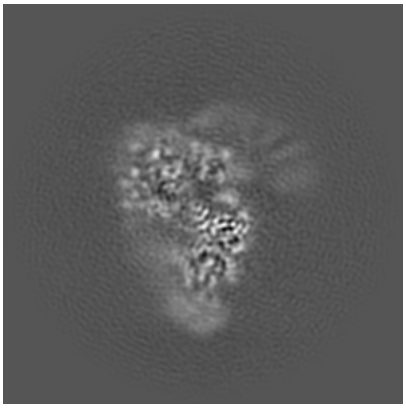


Z

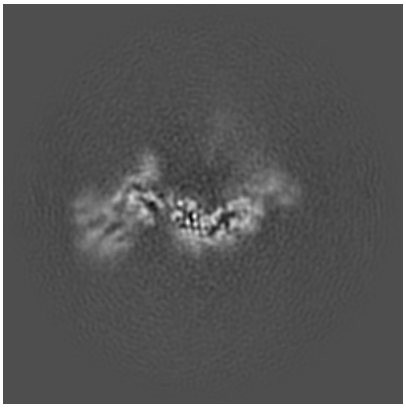
The images above show the map projected in three orthogonal directions.

6.2 Central slices

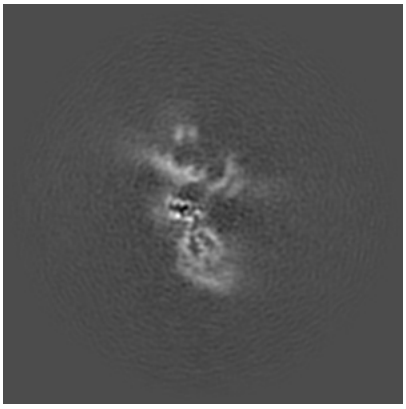
6.2.1 Primary map



X Index: 192



Y Index: 192

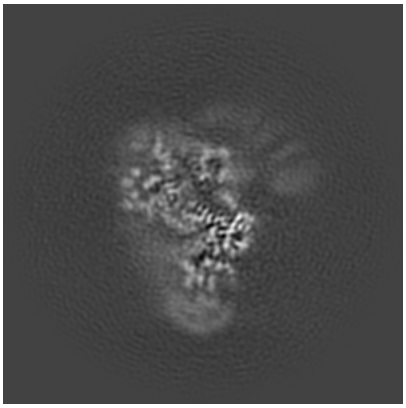


Z Index: 192

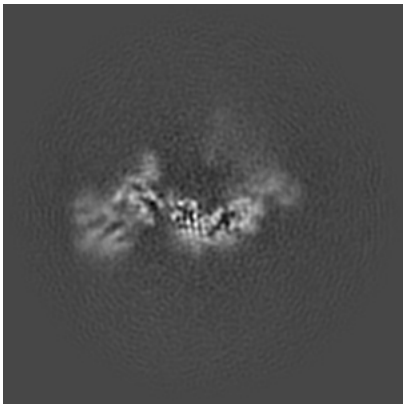
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

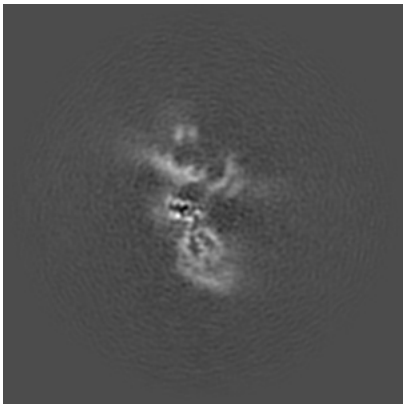
6.3.1 Primary map



X Index: 189



Y Index: 193

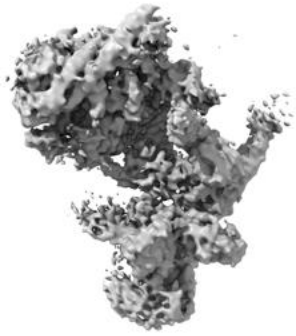


Z Index: 192

The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

6.4.1 Primary map



X

Y

Z

The images above show the 3D surface view of the map at the recommended contour level 0.23. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

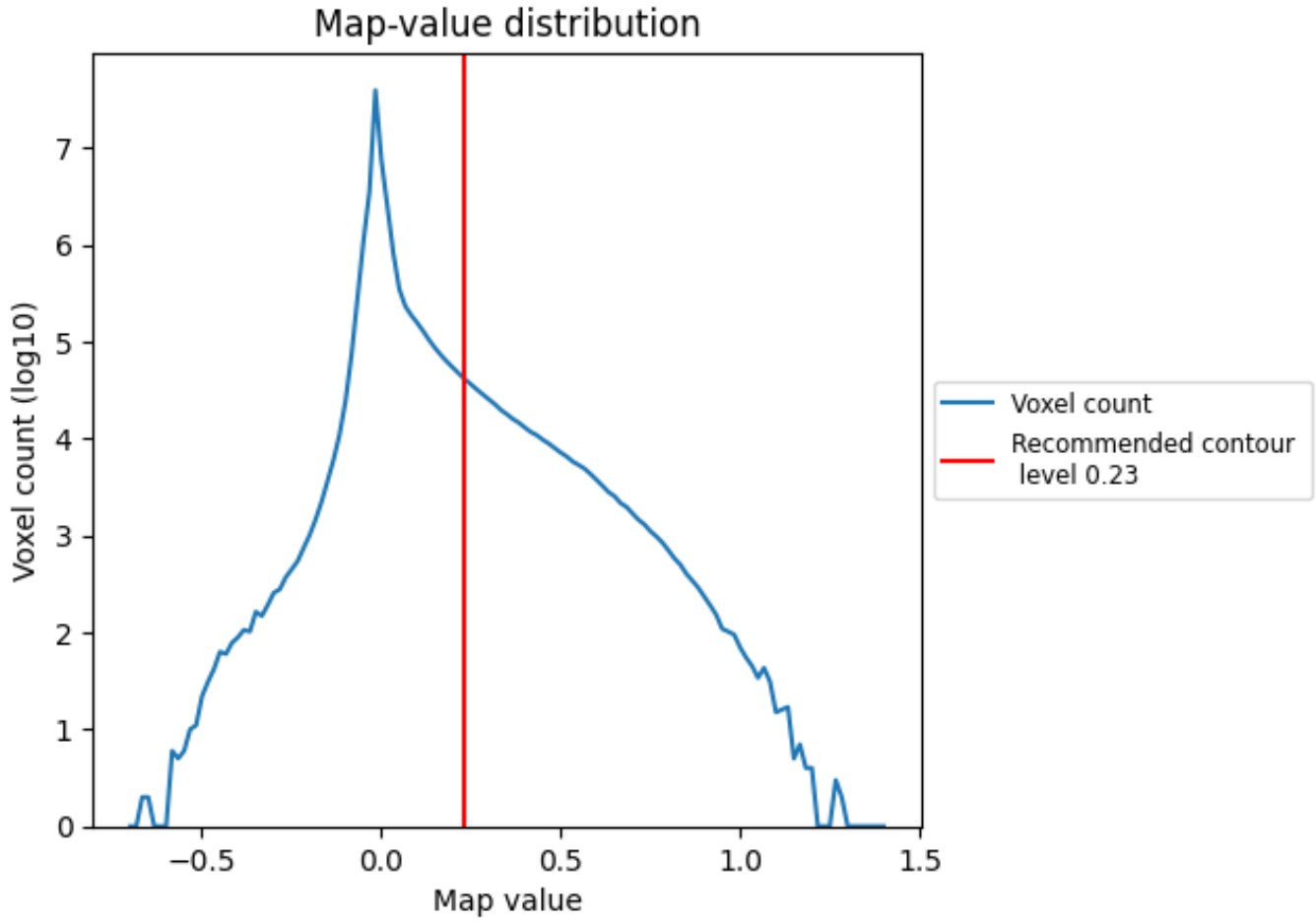
6.5 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

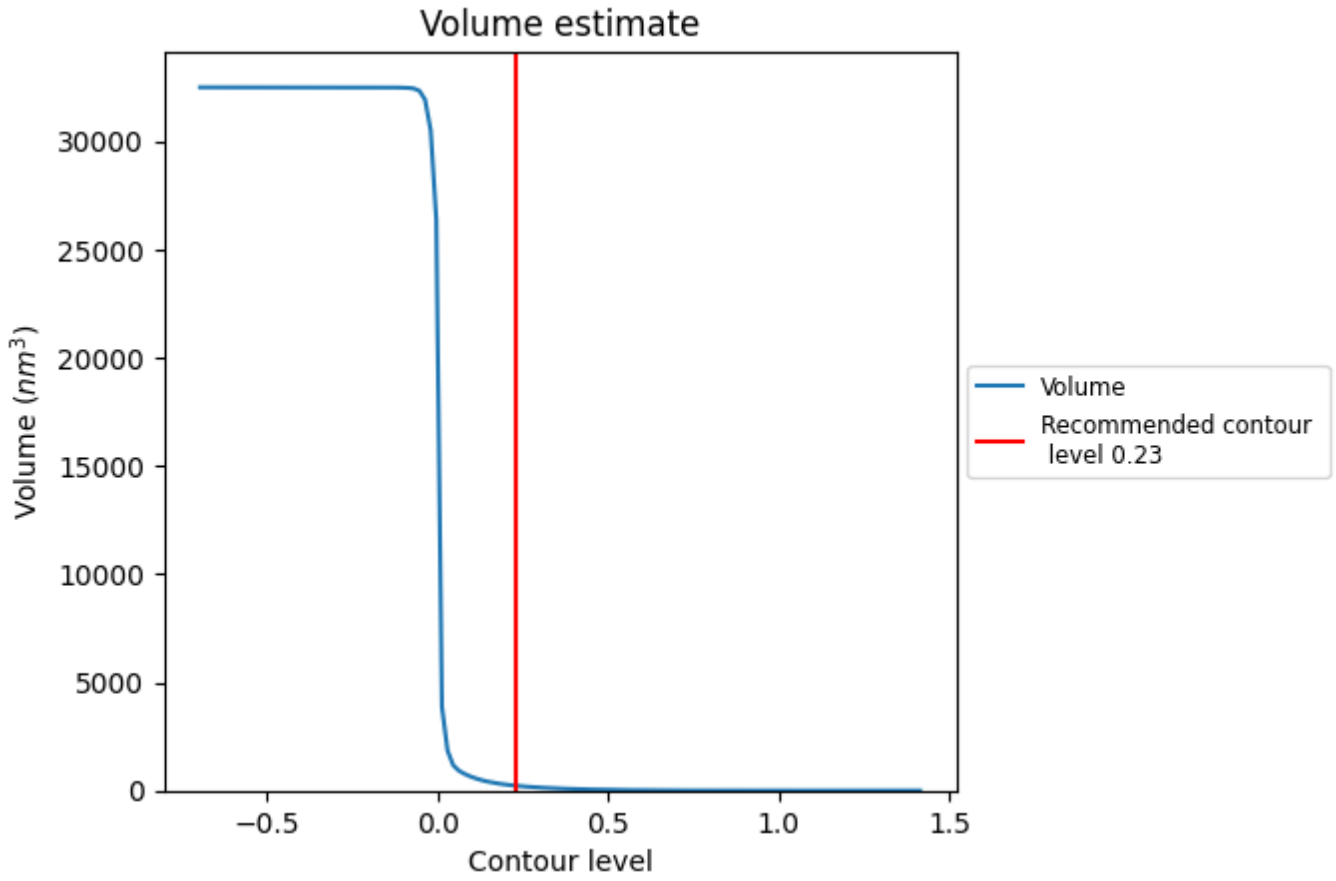
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

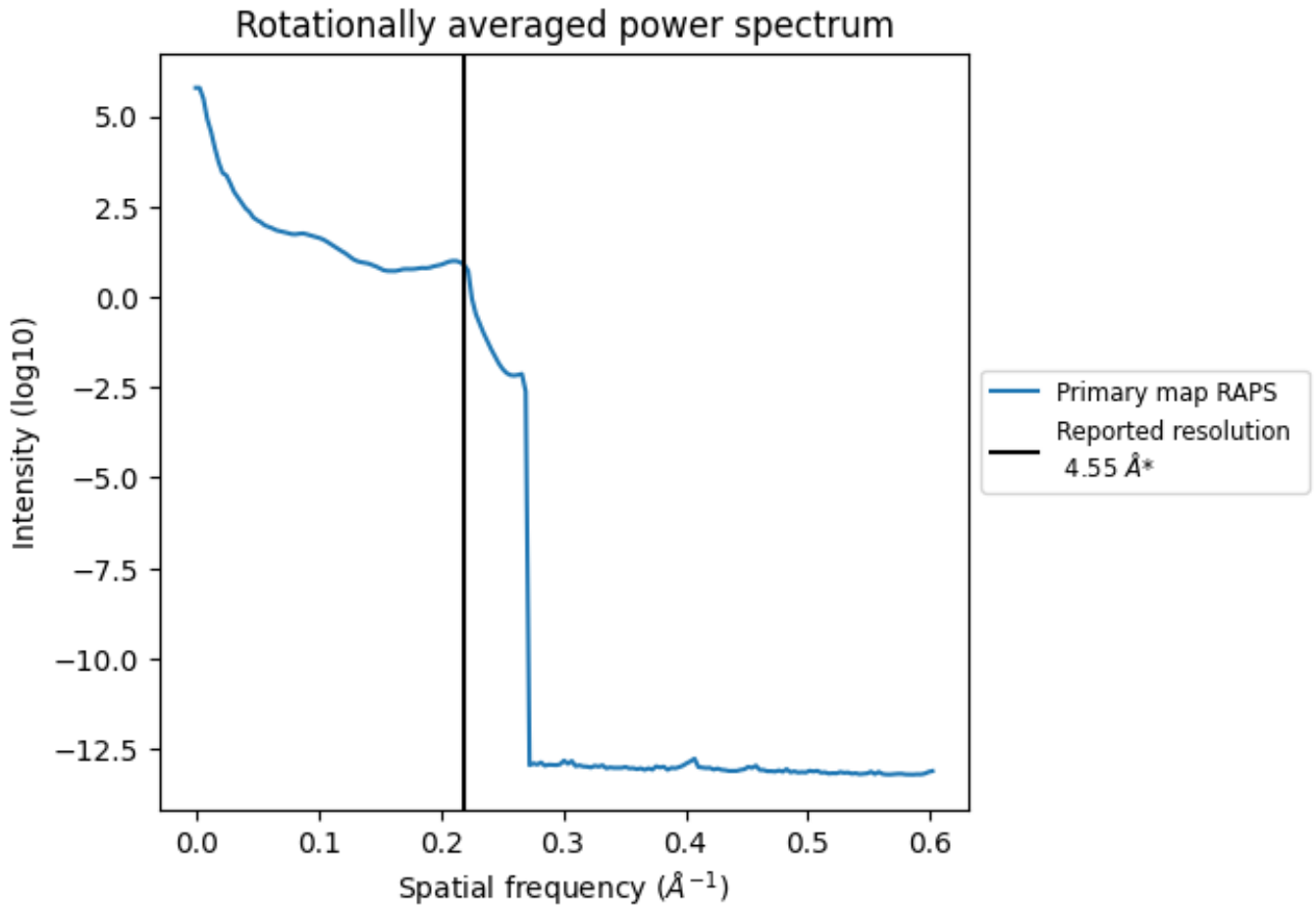
7.2 Volume estimate (i)



The volume at the recommended contour level is 224 nm^3; this corresponds to an approximate mass of 202 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum (i)

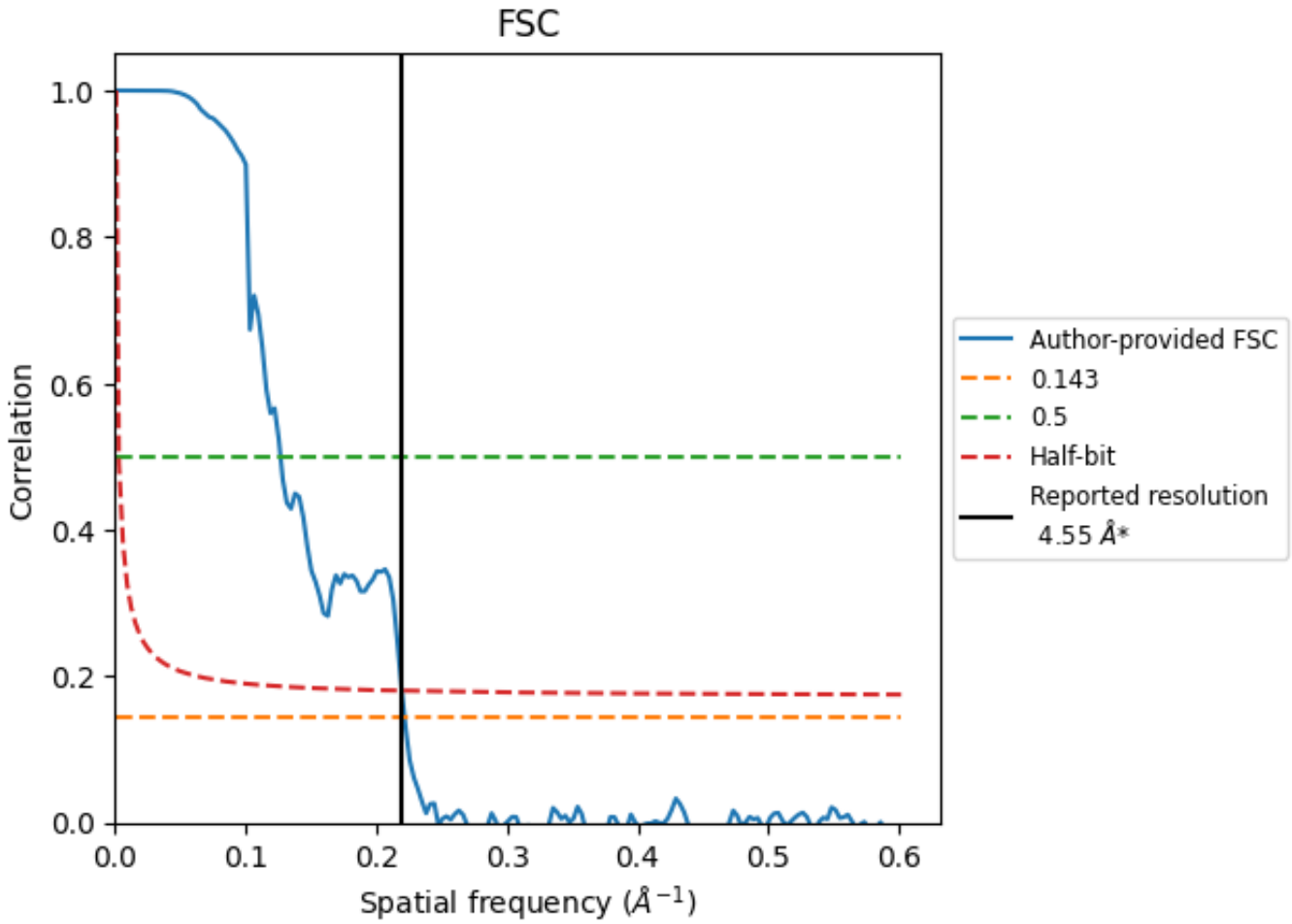


*Reported resolution corresponds to spatial frequency of 0.220 Å^{-1}

8 Fourier-Shell correlation

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.

8.1 FSC



*Reported resolution corresponds to spatial frequency of 0.220 Å^{-1}

8.2 Resolution estimates (i)

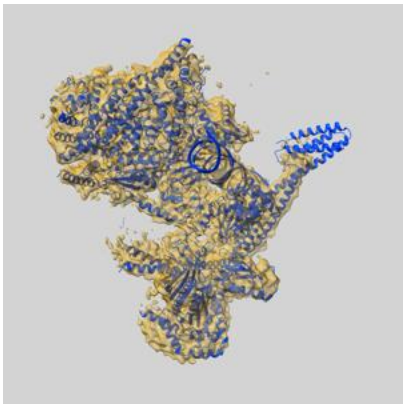
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.55	-	-
Author-provided FSC curve	4.51	7.89	4.56
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

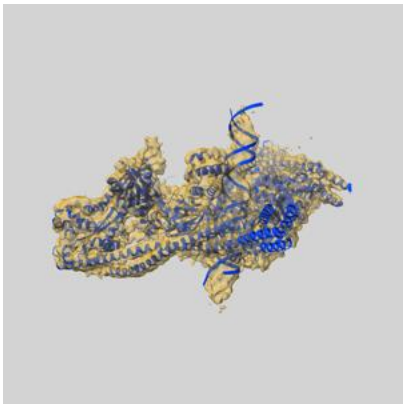
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-14341 and PDB model 7R5V. Per-residue inclusion information can be found in section 3 on page 6.

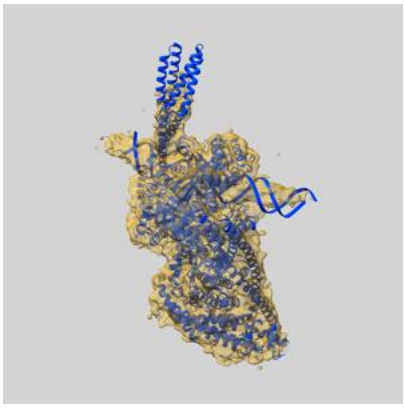
9.1 Map-model overlay



X



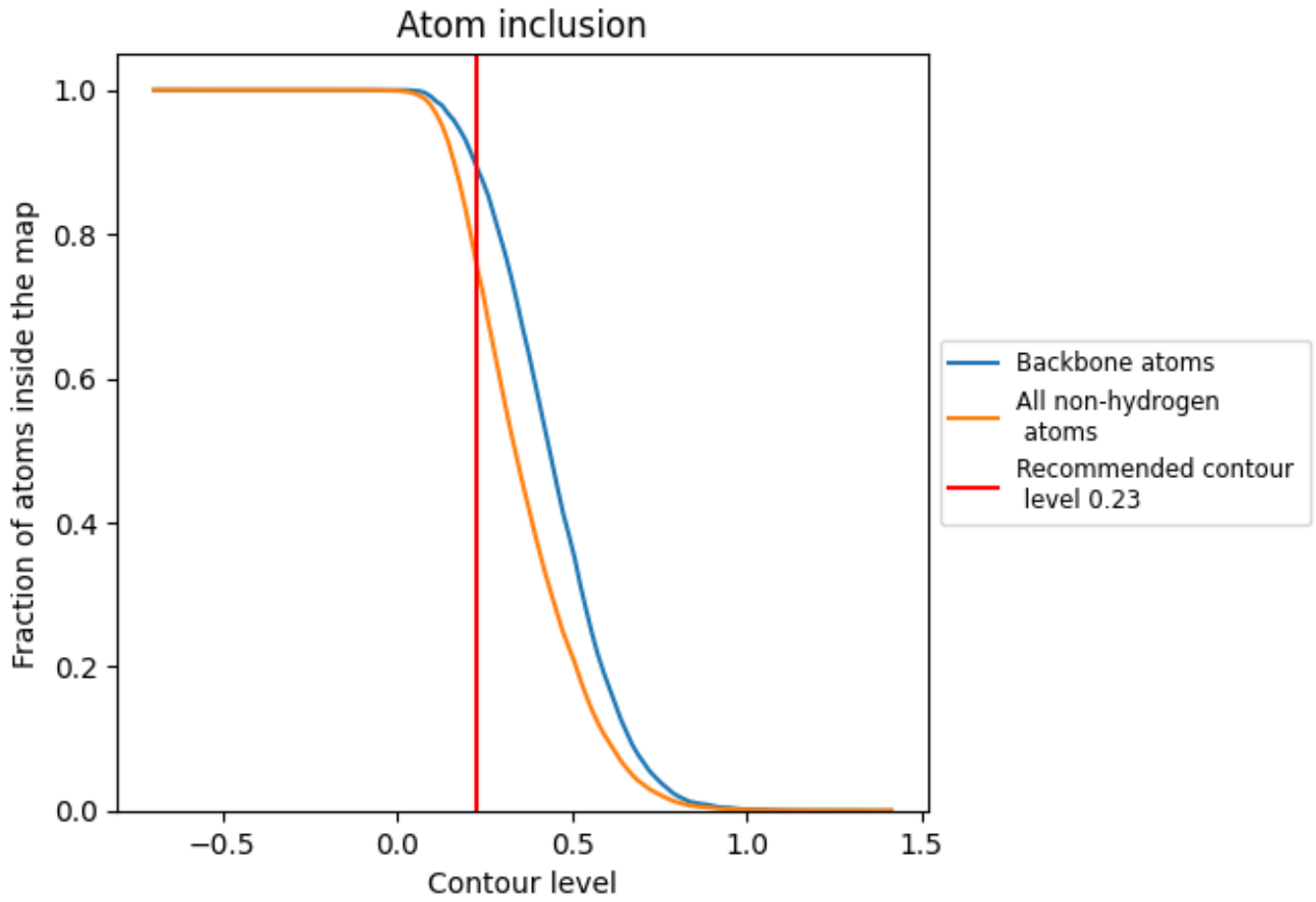
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.23 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Atom inclusion (i)



At the recommended contour level, 89% of all backbone atoms, 75% of all non-hydrogen atoms, are inside the map.